

Analisis Pengaruh Normalisasi, TF-IDF, Pemilihan *Feature-set* Terhadap Klasifikasi Sentimen Menggunakan *Maximum Entropy* (Studi Kasus : Grab dan Gojek)

¹Muhamad Fauzan Putra, ²Anisa Herdiani, ³Diyas Puspandari

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹mfauzanpra@students.telkomuniversity.ac.id, ²anisaherdiani@telkomuniversity.ac.id,

³diyaspuspandari@telkomuniversity.ac.id

Abstrak

Penggunaan sosial media untuk mengutarakan pendapat sudah menjadi kebiasaan masyarakat. Opini yang ditulis konsumen sedikit banyak dapat berpengaruh terhadap bisnis perusahaan. Perusahaan butuh untuk mengevaluasi pelayanan demi kepuasan konsumen dan opini ini dapat sangat bermanfaat dalam berjalannya proses tersebut. Dengan banyaknya *tweet* dari konsumen merupakan pekerjaan yang tidak mudah bagi manusia. Oleh karena itu, pada penelitian ini dibuat suatu sistem yang mampu mengklasifikasi berbagai opini konsumen yang bersifat positif dan negatif. Untuk menentukan fitur dan bobotnya digunakan metode TF-IDF, serta metode Maximum Entropy untuk melakukan klasifikasi. Hasil terbaik yang didapat pada percobaan ini adalah akurasi sebesar 90,67% dan *f1-score* sebesar 84,3%.

Kata kunci: *tweet*, opini, klasifikasi, *tf-idf*, *maximum entropy*

Abstract

The use of social media to express opinions has become a habit of society today. Consumer opinions could more or less affect to the company's business. Companies need to evaluate their service for customer satisfaction and this opinions can be very useful in the process. With the numbers of tweets that must be evaluated from consumers, it isn't an easy task for human. Therefore, in this study, a system has been created that is able to classify consumer opinions into positive and negative classes. Using TF-IDF method to determine the features and weights, and the Maximum Entropy method for classification. The best results obtained in this experiment were accuracy of 90.67% and f1-score of 84.3%.

Keywords: *tweet*, opinion, classification, *tf-idf*, *maximum entropy*

1. Pendahuluan

Penggunaan sosial media untuk mengutarakan pendapat terhadap suatu objek sudah menjadi kebiasaan masyarakat. Salah satunya adalah Twitter yang sejak awal kemunculannya hingga saat ini masih memiliki pengguna setia tak terkecuali di Indonesia yang pada pertengahan 2015 tercatat sebanyak 50 juta pengguna [1].

Di Indonesia beberapa tahun belakangan ini ramai tentang penggunaan jasa transportasi berbasis *online*. Fenomena ini pun tak luput dari pembicaraan masyarakat pada media sosial salah satunya Twitter. Mulai dari berbagi pengalaman, menyampaikan keluhan, dan lain sebagainya. Pada Twitter sendiri perusahaan transportasi *online* memiliki akun resmi yang salah satunya dimanfaatkan sebagai media konsumen untuk memberikan komentar, pendapat, atau kritik. Semua *tweet* dari konsumen tersebut dapat digunakan perusahaan untuk mengevaluasi pelayanan. Mengerti pendapat konsumen merupakan suatu aspek yang penting dalam bisnis [2]. Agar bisnis atau produknya berkembang dengan baik, citra perusahaan harus dikenal baik oleh publik. Maka dari itu opini publik menjadi unsur penting yang harus dikendalikan agar tetap menjadi opini publik yang baik [3].

Namun masalah dalam kasus ini adalah semakin banyaknya suatu *tweet* semakin menyulitkan pihak manusia dalam mengevaluasi opini ataupun pendapat yang ada. Hal ini tentu saja sangat menyulitkan untuk mengekstrak kalimat, membaca, meringkas, lalu mengolahnya menjadi bentuk yang berguna dalam jumlah besar.

Berdasarkan penelitian yang telah dilakukan oleh Aqsath Rasyid Naradipa dan Ayu Purwarianti menyebutkan bahwa mereka menggunakan metode *Maximum Entropy* (ME) dan juga *Support Vector Machine* (SVM) untuk mengklasifikasi sebanyak 150 *review* pada salah satu perusahaan asal Indonesia, didapatkanlah hasil terbaik jika menggunakan metode ME menunjukkan hasil akurasi 86.7% dengan ketentuan data yang digunakan harus formal [4]. Referensi dari Chris Nicolls dan Fei Song diterbitkan pada tahun 2009, dalam penelitiannya mengenai pemilihan *feature set* untuk meningkatkan performa dari *text classification*. Mereka menyebutkan bahwa pemilihan *feature set* pada *sentiment analysis* cukup berpengaruh terhadap performa serta menggunakan metode *maximum entropy* untuk

metode pengklasifikasiannya. Hal ini dibuktikan dari penelitian mereka dengan menggunakan metode POS *Weighting* dengan kombinasi 1, 2, 3, 4, dan 5 untuk kategori *noun*, *verb*, *adv.*, dan *adj* menunjukkan akurasi sebesar 79.41%. Adapun *dataset* yang digunakan sebanyak 574 *reviews* untuk 13 produk yang berbeda dengan perbandingan *review* positif dan negatif 69.5% : 30.5% [5]. Kemudian penelitian yang dilakukan oleh Harlili dan Yudi Wibisono telah melakukan penelitian mengenai *Sentiment Analysis* dengan menggunakan metode *Naive Bayes Classifier* dan untuk *feature* ekstraksinya menggunakan TFxIDF karena teknik tersebut yang paling sederhana menurut mereka. Mereka menyebutkan bahwa performa dalam mengklasifikasi data (*tweet*) mencapai angka 91.3% untuk klasifikasinya sedangkan untuk ekstraksinya hanya menyentuh angka 34.8%, adapun data yang digunakan sebanyak 963 *tweet* [6].

Oleh karena itu diperlukan adanya suatu sistem yang dapat mengotomatisasi pekerjaan tersebut. Dengan menggunakan analisis sentimen yang mana merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung pada suatu kalimat baik yang bernilai positif, negatif, maupun netral [7]. Menggunakan metode tf-idf untuk memberikan bobot setiap fitur dan metode *Maximum Entropy* untuk klasifikasinya.

Pada penelitian ini digunakan *tweet* berbahasa Indonesia yang didapat dari akun resmi GO-JEK dan GRAB Indonesia yang mayoritas isi dari *tweet* akan menggunakan bahasa sehari-hari yang jauh dari ketentuan KBBI. Kemudian data tersebut dilabeli secara manual.

Penelitian ini bertujuan untuk membangun suatu sistem yang dapat mengklasifikasi secara otomatis suatu *tweet* menjadi kelas positif, negatif atau netral. Pada prosesnya digunakan proses normalisasi yaitu proses dimana kata yang terkandung pada *tweet* akan disesuaikan dengan ketentuan KBBI. Selanjutnya *tweet* tersebut akan dihitung bobot masing-masing katanya dengan metode TF-IDF untuk dapat memilih fitur utama dalam suatu *tweet*. Kemudian dalam suatu kalimat terdapat berbagai jenis kata yang terkandung (*Part of Speech*) yang tentunya setiap jenis kata punya pengaruh yang berbeda terhadap nilai suatu kalimat. Dalam penelitian ini penulis ingin mengetahui jenis kata apa yang paling berpengaruh terhadap proses klasifikasi opini dalam kasus ini sehingga dapat ditentukan bobot setiap jenis kata.

2. Studi Terkait

2.1 Analisis Sentimen

Analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen [7].

2.2 Klasifikasi Maximum Entropy

Entropy sendiri merupakan nilai yang tidak pasti atau yang lebih dikenal sebagai nilai probabilitas. Sedangkan *maximum entropy* yaitu model untuk mencari nilai tertinggi dari suatu probabilitas atau *entropy*. Probabilitas dalam kasus ini merupakan prediksi informasi yang ada pada suatu *tweet*. Dengan menggunakan *maximum entropy* akan dihitung probabilitas informasi dari suatu *tweet*, yang kemudian akan dikelompokkan menjadi bernilai positif atau negatif. Fungsi umum dari *entropy* adalah sebagaimana pada persamaan (1) berikut

$$entropy(S) = - \sum_{i=1}^n P(S_i) \log_2 P(S_i) \quad (1)$$

Maximum entropy secara definisi merupakan metode yang digunakan untuk mencari probabilitas distribusi yang dimana probabilitas tersebut memiliki nilai *entropy* tertinggi. Teknik ini juga merupakan salah satu metode yang menggunakan statistika [7]. Dalam arti lain, *Maximum Entropy* adalah suatu teknik untuk mempelajari dan

mengetahui suatu probabilitas atau peluang distribusi dari suatu data tertentu, teknik ini akan mencari distribusi peluang dari suatu kejadian berdasarkan suatu konteks probabilitas yang seragam [8] [9].

Dalam Tugas Akhir ini penyusun mendefinisikan bahwa hasil adalah anggota dari suatu himpunan (y) dan kondisi yang mempengaruhi nilai hasil tersebut didefinisikan sebagai konteks, dimana konteks ini merupakan anggota dari himpunan (x). Himpunan (x) ini adalah kumpulan *feature* yang digunakan dalam Tugas Akhir ini. Hasil yang akan didapatkan dengan menggunakan teknik ini adalah probabilitas himpunan (y) terhadap (x) hal ini dipresentasikan ke dalam fungsi $f(x,y)$ dan prediksi dari faktor yang mempengaruhi dipresentasikan sebagai $p(x)$. Fungsi tersebut dapat dilihat pada persamaan (2) berikut ini

$$f(x,y) = \begin{cases} 1 & , \text{jika } y' = y \text{ dan } cp(x) = true \\ 0 & , \text{lainnya} \end{cases} \quad (2)$$

Dimana $cp(x)$ sebagai conditional probabilitas suatu keadaan (y) terhadap (x) dapat dilihat dari fungsi pada persamaan (3) berikut ini:

$$p(y|x) = \frac{\prod_i a_i^{f_i(x,y)}}{Z_a(x)} \quad (3)$$

Dengan ketentuan Z adalah nilai normalisasi dari setiap kata yang terdapat dalam himpunan (x). Dimana nilai Z, dapat dilihat pada persamaan (4) berikut ini:

$$Z_a(x) = \sum_y \prod_i a_i^{f_i(x,y)} \quad (4)$$

2.3 Model TF-IDF

Sebelum diklasifikasi, suatu *tweet* perlu diketahui topik yang sedang dibahas (*feature*). Dengan menggunakan model *term frequency-inverse document frequency* (TF-IDF) dapat diketahui bahasan apa yang sedang dikemukakan pada suatu *tweet*. Selain itu pada suatu kalimat (*tweet*) terdapat kata subjek, objek, kata kerja dan sebagainya yang tentu saja pasti berbeda nilai kepentingan dalam suatu kalimat (bobot). Dengan menggunakan teknik TF-IDF ini pun dapat menentukan bobot dari setiap kata sehingga kata yang berbobot paling tinggi dapat diasumsikan sebagai topik utama dalam suatu teks atau kalimat. Fungsi dari perhitungan bobot dari suatu kata digambarkan melalui persamaan (5) berikut.

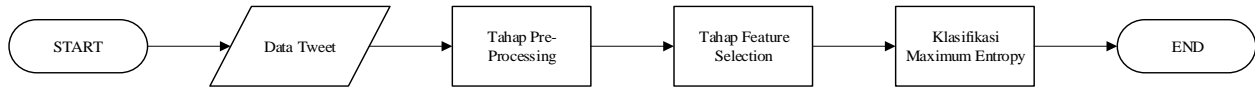
$$f_{w,c^i}(d,c) = \begin{cases} N(d,w)/N(d), & \text{jika } c = c' \\ 0, & \text{lainnya} \end{cases} \quad (5)$$

Penggunaan metode ini membutuhkan definisi awal sebagai strength dari masing-masing tags yang telah disebutkan sebelumnya dengan komposisi nilai [0,5] dan untuk mencari fungsi normal dari POS strength ini didefinisikan pada persamaan (6) berikut ini

$$f_{w,c'}(d,c) = \begin{cases} str(POS_w)/\sum_{i \in d} str(POS_i), & \text{jika } c = c' \\ 0, & \text{lainnya} \end{cases} \quad (6)$$

$N(d,w)$ = Sebagai nilai kemunculan text (w) pada review (d)
 $N(d)$ = Sebagai total kata yang terdapat dalam review (d)
 $POS_{(n)}$ = Sebagai nilai Tags yang telah ditentukan
 $f_{w,c'}(d,c)$ = Sebagai nilai bobot dari suatu feature tertentu

3. Perancangan Sistem



Gambar 3-1. Gambaran umum sistem

3.1 Tahap Pre-processing

3.1.1 Case Folding

Mengubah huruf besar (kapital) yang terdapat pada tweet dilihat pada Tabel 3-1.

Tabel 3-1 Contoh *Case Folding*

Input
Sorry nih kesabaran gw udh habis, gw udah 5 kali merasa dirugikan @grabid https://t.co/hanxmyco7r
Output
sorry nih kesabaran gw udh habis, gw udah 5 kali merasa dirugikan @grabid https://t.co/hanxmyco7r

3.1.2 Remove URL

Menghapus URL yang terdapat pada *tweet* dilihat pada Tabel 3-2.

Tabel 3-2 Contoh *Remove URL*

Input
sorry nih kesabaran gw udh habis, gw udah 5 kali merasa dirugikan @grabid https://t.co/hanxmyco7r
Output
sorry nih kesabaran gw udh habis, gw udah 5 kali merasa dirugikan @grabid

3.1.3 Clean Number & Unused Character

Proses menghilangkan angka dan tanda baca pada *tweet* dilihat pada Tabel 3-3.

Tabel 3-3 Contoh *Clean Number & Unused Character*

Input
sorry nih kesabaran gw udh habis, gw udah 5 kali merasa dirugikan @grabid
Output
sorry nih kesabaran gw udh habis gw udah kali merasa dirugikan @grabid

3.1.4 Stopword Removal & Normalisasi

Proses mengubah kata menjadi kata baku dan mengecek jika kata tersebut merupakan *stopword* akan dihapus. Kemudian mengubah kata menjadi kata dasar. Dapat dilihat pada Tabel 3-4.

Tabel 3-4 Contoh *Stopword Removal & Normalisasi*

Input
sorry nih kesabaran gw udh habis gw udah kali merasa dirugikan @grabid
Output
maaf sabar habis rasa rugi <brand_2>

3.1.5 Tokenisasi & POS Tagging

Mengubah seluruh kata menjadi token dan memberikan tanda jenis kata dilihat pada Tabel 3-5.

Tabel 3-5 Contoh Tokenisasi & POS Tagging

Input		
maaf sabar habis rasa rugi <brand_2>		
Output		
Index	Word	Tag
1	maaf	n
2	sabar	adj
3	habis	v
4	rasa	n
5	rugi	adj
6	<brand_2>	<?>

3.2 Feature Selection (TF-IDF)

Setelah token dari semua *tweet* selanjutnya menilai bobot dari setiap token tersebut. Dapat dilihat contoh pada Tabel 3-6.

Tabel 3-6 Contoh proses perhitungan TF-IDF

No.	Word	Size	TF	Total	IDF	TF.IDF
1	maaf	1	0.77	29	3.197	0.246
2	sabar	1	0.77	2	5.832	0.449
3	habis	1	0.77	4	5.142	0.396
4	rasa	1	0.77	8	4.454	0.343
5	rugi	1	0.77	31	3.133	0.241
6	<brand_2>	1	0.77	270	1.258	0.097

3.3 Klasifikasi (Maximum Entropy)

Setelah didapatkan nilai TF-IDF kemudian dilakukan klasifikasi terhadap *tweet* untuk menentukan kelas. Dapat dilihat contoh pada Tabel 3-7.

Tabel 3-7 Contoh *tweet* yang telah diklasifikasi

Tweet ID	Tweet	Features	Class
1001	Sorry nih kesabaran gw udh habis, gw udah 5 kali merasa dirugikan @grabid https://t.co/hanxmyco7r	Maaf, Sabar, Rugi	N
1002	Puas banget pakai grabfood banyak promonya, mantap @grab	Puas, Mantap	P
...

Sebagai contoh perhitungan jika suatu *tweet* berisi “Pelayanan customer service jelek, mengecewakan”. Dengan model ini dicari dahulu probabilitas setiap kata utama pada *tweet* terhadap setiap kelas. Didapatkanlah nilai probabilitas untuk kata “layan” untuk kelas positif 0.2, untuk kelas negatif 0.75, dan kelas netral 0.05. Kemudian kata “jelek” untuk kelas positif 0, kelas negatif 0.95, kelas netral 0.05. Kemudian kata “kecewa” untuk kelas positif 0.01, kelas negatif 0.9, dan kelas netral 0.09. Selanjutnya dari proses *training* telah didapatkan bobot setiap kata, yaitu:

$$w(\text{layan}) = -3.32$$

$$w(\text{jelek}) = -3.51$$

$$w(\text{kecewa}) = -9.5$$

$$P(1|t1) = 0.2 \quad P(2|t1) = 0.75 \quad P(3|t1) = 0.05$$

$$P(1|t2) = 0 \quad P(2|t2) = 0.95 \quad P(3|t2) = 0.05$$

$$P(1|t3) = 0.01 \quad P(2|t2) = 0.9 \quad P(3|t3) = 0.09$$

$$p(1 | Di) = || (-3.32 \times 0.20) + (-3.51 \times 0.00) + (-9.5 \times 0.05) || = 0.974$$

$$p(2 | Di) = || (-3.32 \times 0.75) + (-3.51 \times 0.95) + (-9.5 \times 0.90) || = 14.21$$

$$p(3 | Di) = || (-3.32 \times 0.05) + (-3.51 \times 0.05) + (-9.5 \times 0.09) || = 0.073$$

Nilai untuk *tweet* tersebut dilakukan normalisasi, dapat dicari sebagai berikut

$$P(1 | Di) = \frac{0.974}{0.974 + 14.21 + 0.073} = 0.064$$

$$P(2 | Di) = \frac{14.21}{0.974 + 14.21 + 0.073} = 0.931$$

$$P(3 | Di) = \frac{0.073}{0.974 + 14.21 + 0.073} = 0.005$$

Dari perhitungan diatas dapat diambil kesimpulan bahwa *tweet* “Pelayanan customer service jelek, mengecewakan” diprediksi bernilai negatif dengan nilai probabilitas sebesar 93,1%.

3.4 Evaluasi

Evaluasi dari model sistem yang dibangun pada tugas akhir ini menggunakan parameter benar atau tidaknya sistem dalam memprediksi hasil yang didapatkan, hal ini perlu digaris bawahi bahwasanya pengukuran performa pada tugas akhir ini dengan menggunakan metode *confusion matrix* yaitu metode yang digunakan untuk menilai performa dari kasus sentimen analisis yang berbentuk *binomial* dengan memiliki 4 nilai patokan [10]. Metode ini dipilih karena teknik pengklasifikasian pada tugas akhir ini menggunakan *maximum entropy* yang menggunakan peluang atau tidak pasti.

Dalam tugas akhir ini terdapat tiga kelas yaitu Positif, Negatif, dan Netral, oleh karena itu evaluasi dilakukan terhadap masing-masing kelas kemudian ketiga nilai dapat digabungkan. Adapun 4 *rules* nilai sebagai berikut:

- True Positive (TP), apabila hasil prediksi bernilai P dan data yang digunakan bernilai P.
- True Negative (TN), apabila hasil prediksi bernilai ~P dan data yang digunakan bernilai ~P.
- False Positive (FP), apabila hasil prediksi bernilai P dan data yang digunakan bernilai ~P.
- False Negative (FN), apabila hasil prediksi bernilai ~P dan data yang digunakan bernilai P.

Rules diatas berlaku untuk kelas Positif, untuk kedua kelas lainnya dapat diganti P menjadi N dan Neu.

Dapat dilihat dari Tabel 3-1 berikut ini.

Tabel 1. *Confusion Matrix*

Data \ Prediksi	1	0
1	TP	FP
0	FN	TN

Beberapa patokan tersebut dapat dijadikan parameter dalam penentuan performa dari model klasifikasi yang dibangun. Diantaranya yaitu:

- *Accuracy* adalah perhitungan seberapa banyak persentase prediksi klasifikasi yang sesuai dan akurat dari keseluruhan *tweet* yang diberikan, dengan persamaan (7) sebagai berikut:

$$A = \frac{TP + TN}{\text{jumlah tweet}} \quad (7)$$

- *Precision* adalah perhitungan seberapa banyak persentase prediksi klasifikasi yang sesuai dan juga bernilai opini positif (1) dibandingkan dari jumlah keseluruhan prediksi yang bernilai opini positif (1), dengan persamaan (8) sebagai berikut:

$$P = \frac{TP}{TP + FP} \quad (8)$$

- *Recall* adalah perhitungan seberapa banyak persentase prediksi klasifikasi yang sesuai dan juga bernilai opini positif (1) dibandingkan dari jumlah keseluruhan *tweet* yang bernilai opini positif (1), dengan persamaan (9) sebagai berikut:

$$R = \frac{TP}{TP + FN} \quad (9)$$

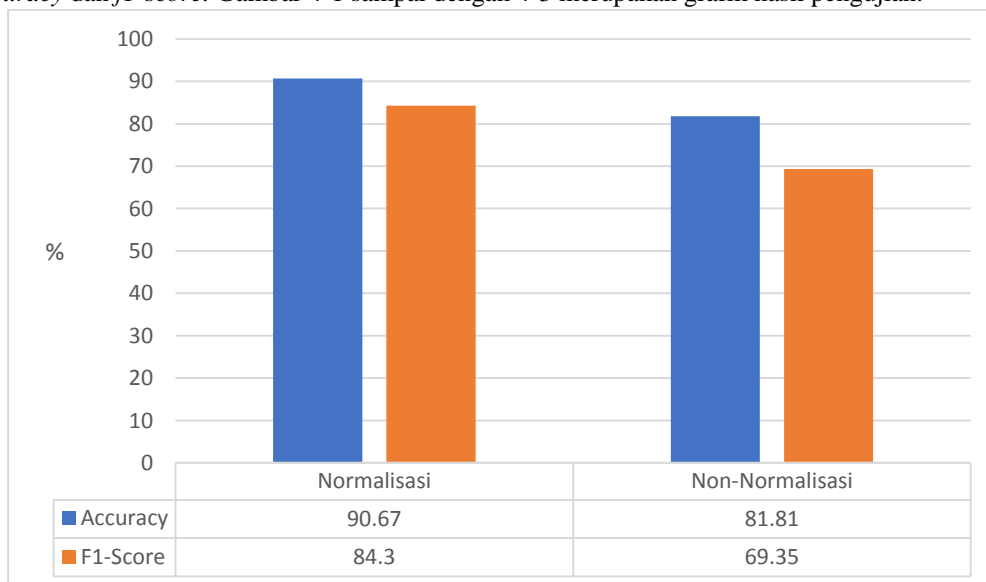
- F1-Measure adalah nilai yang dapat mewakili keseluruhan kinerja sistem dan merupakan penggabungan antara nilai *precision* dan nilai *recall*, dengan persamaan (10) sebagai berikut:

$$F1 - \text{measure} = 2 \times \frac{P \times R}{P + R} \quad (10)$$

4. Evaluasi Sistem

Proses klasifikasi dalam sistem ini menggunakan metode *Maximum Entropy* yang memberikan kelas pada suatu *tweet* menjadi kelas positif, negatif atau bukan keduanya (netral). Sistem telah melalui 100 kali iterasi proses *training* dengan menggunakan 1000 data *tweet*. Setelah melalui proses *training* barulah sistem akan dilakukan proses evaluasi. Untuk mendapatkan hasil yang optimal maka sistem akan melalui beberapa skema pengujian. Data yang digunakan untuk tahap testing sebesar 700 data dengan komposisi 94 *tweet* bernilai positif, 461 bernilai negatif, dan 145 bernilai netral.

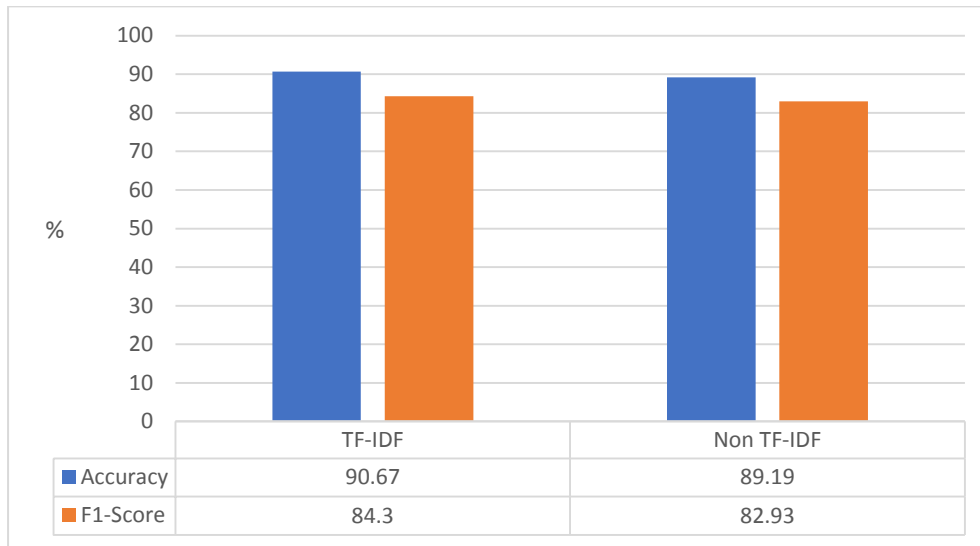
Dalam penelitian ini akan dilakukan 3 jenis pengujian, yaitu; Pengaruh tahap normalisasi terhadap hasil klasifikasi, pengaruh perhitungan TF-IDF terhadap hasil klasifikasi, dan pengaruh penggunaan jenis *feature-set* (*adjective*, *conjunction*, *verb*, *adverb*, dan *noun*) terhadap hasil klasifikasi. Hasil dalam pengujian ini diukur dengan nilai *accuracy* dan *f1-score*. Gambar 4-1 sampai dengan 4-3 merupakan grafik hasil pengujian.



Gambar 4-1 Hasil evaluasi pengaruh penggunaan Normalisasi

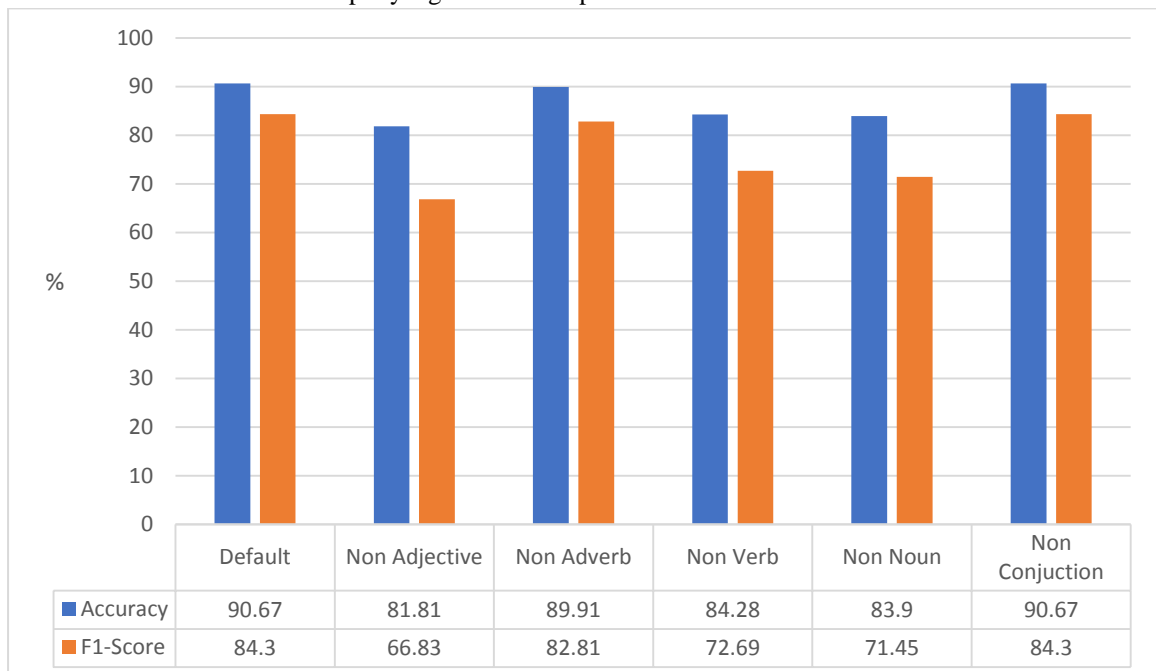
Pada Gambar 4-1 terdapat grafik yang menunjukkan hasil evaluasi yang menggunakan tahap Normalisasi dengan yang tidak menggunakan Normalisasi. Sesuai grafik tersebut dapat dilihat bahwa penggunaan Normalisasi dapat mempengaruhi hasil klasifikasi dengan selisih akurasi sebesar 8,86% dan *f1-score* sebesar 14,95%. Penggunaan

Normalisasi merupakan hal yang penting mengingat dalam *tweet* berbahasa Indonesia terdapat banyak sekali penggunaan kata-kata tidak baku sehingga sistem tidak akan dapat mendeteksi kata tersebut untuk dilakukan proses klasifikasi.



Gambar 4-2 Hasil evaluasi pengaruh penggunaan TF-IDF

Pada Gambar 4-2 terdapat grafik yang menunjukkan hasil evaluasi antara yang menggunakan perhitungan TF-IDF dengan yang tidak menggunakan TF-IDF. Dari grafik tersebut dapat dilihat bahwa penggunaan perhitungan TF-IDF cukup mempengaruhi performansi dari proses klasifikasi dengan selisih akurasi sebesar 1,48% dan *f1-score* sebesar 1,37%. TF-IDF sendiri memiliki peran penting dalam proses klasifikasi, memungkinkan sistem untuk tidak memberikan bobot yang terlalu besar untuk suatu kata yang sering muncul dalam dokumen sehingga dapat membuat sistem lebih akurat dalam menentukan topik yang dibicarakan pada suatu *tweet*.



Gambar 4-3 Hasil evaluasi pengaruh setiap jenis kata

Pada Gambar 4-3 terdapat grafik yang menggambarkan hasil evaluasi dari sistem terhadap jenis kata. Jenis kata yang digunakan antara *Adjective* (kata sifat), *Adverb* (kata keterangan), *Verb* (kata kerja), *Noun* (kata benda), dan *Conjunction* (kata hubung). Dilihat dari grafik bahwa jenis kata yang paling berpengaruh dalam proses klasifikasi

adalah kata sifat yang jika tanpa disertakan kata sifat pada kamus akan mengalami penurunan akurasi sebesar 8,86% dan *f1-score* sebesar 17,47% sehingga menjadi jenis kata paling berpengaruh terhadap hasil klasifikasi. Sebaliknya kata hubung merupakan jenis kata yang paling tidak memiliki peran dalam proses klasifikasi. *Conjunction* atau kata hubung sebgiaan besar merupakan *stopword* sehingga tidak akan berpengaruh terhadap pengklasifikasian, sedangkan kata *conjunction* yang bukan merupakan *stopword* tidak dapat ditemukan pada *dataset*, oleh karena itu pada penelitian ini jenis kata tersebut sama sekali tidak mempengaruhi hasil evaluasi karena pada dasarnya seluruh kata *conjunction* dianggap tidak ada. Sebaliknya untuk kata *adjective* dan *noun* sebagian besarnya bukan merupakan *stopword*.

5. Kesimpulan dan Saran

Berdasarkan hasil pengujian dan analisis yang telah dilakukan sebelumnya, maka dapat diambil kesimpulan dan saran sebagai berikut:

5.1 Kesimpulan

1. Sistem yang dibangun dengan menggunakan metode *Maximum Entropy* ini cukup baik dalam mengklasifikasikan *tweet* menjadi kelas positif, negatif, dan netral dengan hasil akurasi sebesar 90,67% dan *f1-score* sebesar 84,3%.
2. Penggunaan normalisasi atau proses mengubah kata menjadi kata baku yang sesuai KBBI dalam proses klasifikasi cukup mempengaruhi dikarenakan apabila kata tidak diubah menjadi kata baku, kata tersebut tidak akan dihitung oleh sistem.
3. Penggunaan TF-IDF dalam proses klasifikasi ini rupanya hanya sedikit mempengaruhi hasil, namun tetap dalam penggunaannya dapat meningkatkan performa dari sistem.
4. Jenis kata yang paling berpengaruh dalam proses klasifikasi opini ialah kata sifat (*adjective*) dikarenakan pada umumnya seseorang mengeluarkan opini mengandung kata sifat, kemudian kata kerja, dan diikuti oleh kata benda sebagai objek.

5.2 Saran

1. Untuk pekerjaan selanjutnya alangkah baiknya jika penggunaan *part of speech* dilakukan lebih mendalam daripada hanya untuk mempengaruhi *dataset* agar hasil klasifikasi dapat menghasilkan performansi yang lebih baik lagi.
2. Untuk proses pelabelan data secara manual sebaiknya sejak awal dilakukan dibawah pengawasan ahli sehingga dapat menghemat waktu dalam proses evaluasi data.
3. Proses pengembangan sistem klasifikasi perlu memperhatikan spesifikasi dari peranti pendukung pekerjaan, sehingga perlu dipastikan lingkungan sistem sudah memenuhi untuk pengembangan sistem dikarenakan akan mempengaruhi ke durasi pengerjaan.
4. Selanjutnya bisa digunakan pembagian kelompok kata selain unigram sehingga dapat menilai kata negasi seperti “tidak pernah”, “belum baik”, dan lain sebagainya secara akurat ketimbang menggunakan unigram yang hanya menghitung setiap kata tanpa peduli kata di sebelum atau setelahnya yang akan menjadi kesalahan pengartian terhadap inti kalimat.

DAFTAR PUSTAKA

- [1] Maulana. A, "Twitter Rahasiakan Jumlah Pengguna di Indonesia", CNN Indonesia, 23 Maret 2016. [Online]. Available: <https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia> 020318 [Diakses 24 Februari 2018]
- [2] S. Wardoyo, "Opinion Summarization Fitur Produk Elektronik Pada Amazon.com Dengan Metode Maximum Entropy," pp. 7-9, 2012
- [3] Rahmah, M. N. Opini publik dalam pencitraan bisnis. [Online] Available: https://www.academia.edu/9019827/Opini_publik_dalam_pencitraan_bisnis [Diakses 1 Maret 2018]
- [4] A. RN dan A. P, "Sentiment Classification for Indonesian Message in Social Media," dalam Bandung Institute of Tehcnology Conference, Bandung, 2011.
- [5] C. Nicholls and F. Song, "Improving Sentiment Analysis with Part Of Speech Weighting," in *Eighth International Conference on Machine Learning and Cybernetics*, Baoding, 2009.
- [6] Harlili dan Y. Wibisono, "SISTEM ANALISIS OPINI MICROBLOGGING BERBAHASA INDONESIA," dalam Konferensi Nasional Sistem Informasi, Bali, 2012.
- [7] I. F. Rozi, S. Hadi, and E. Achmad, "Implementasi Opinion Mining (AnalisisSentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," vol. 6, no. 1, pp. 37–43, 2012.
- [8] M. C dan S. H, "Foundation of Statistical Natural Language Processing," dalam The International Massachusetts Institute of Technology, USA, 2002.
- [9] D. M. J, "Information Theory, Inference and Learning Algorithms," pp. 3-4, 2003.
- [10] D. M. Powers, "Evaluation From Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation," 2007.
- [11] V. Chandani, F. I. Komputer, and U. D. Nuswantoro, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film," vol. 1, no. 1, pp. 56–60, 2015.
- [12] L. Tylor dan B. Francis, "Reproducible Research With R and R Studio," dalam The Origin of R, T&F Group, 2013, pp. 3-4.